Compatible interlaced SDTV and progressive HDTV

FIELD OF THE INVENTION

The invention relates to a video encoder/decoder, and more particularly to a compatible interlaced SDTV and progressive high resolution low bit rate coding scheme for use by a video encoder/decoder.

5

BACKGROUND OF THE INVENTION

Because of the massive amounts of data inherent in digital video, the transmission of full-motion, high-definition digital video signals is a significant problem in the development of high-definition television. More particularly, each digital image frame is.

10 a still image formed from an array of pixels according to the display resolution of a particular system. As a result, the amounts of raw digital information included in high-resolution video sequences are massive. In order to reduce the amount of data that must be sent, compression schemes are used to compress the data. Various video compression standards or processes have been established, including, MPEG-2, MPEG-4, and H.263.

15 Many applications are enabled where video is available at various resolutions and/or qualities in one stream. Methods to accomplish this are loosely referred to as scalability techniques. There are three axes on which one can deploy scalability. The first is scalability on the time axis, often referred to as temporal scalability. Secondly, there is scalability on the quality axis (quantization), often referred to as signal-to-noise (SNR)

20 scalability or fine-grain scalability. The third axis is the resolution axis (number of pixels in image) often referred to as spatial scalability. In layered coding, the bitstream is divided into two or more bitstreams, or layers. Each layer can be combined to form a single high quality signal. For example, the base layer may provide a lower quality video signal, while the enhancement layer provides additional information that can enhance the base layer image.

25 In particular, spatial scalability can provide compatibility between different video standards or decoder capabilities. With spatial scalability, the base layer video may have a lower resolution than the input video sequence, in which case the enhancement layer carries information which can restore the resolution of the base layer to the input sequence level.

2

Figure 1 illustrates a known spatial scalable video encoder. The depicted encoding system accomplishes layer compression, whereby a portion of the channel is used for providing a low resolution base layer and the remaining portion is used for transmitting edge enhancement information, whereby the two signals may be recombined to bring the

5   system up to high-resolution. The high resolution video input is split by splitter 102 whereby the data is sent to a low pass filter 104 and a subtraction circuit 106. The low pass filter 104 reduces the resolution of the video data, which is then fed to a base encoder 108. In general, low pass filters and encoders are well known in the art and are not described in detail herein for purposes of simplicity. The encoder 108 produces a lower resolution base stream which

10   can be broadcast, received and via a decoder, displayed as is, although the base stream does not provide a resolution which would be considered as high-definition.

The output of the encoder 108 is also fed to a decoder 112 within the system 100. From there, the decoded signal is fed into an interpolate and upsample circuit 114. In general, the interpolate and upsample circuit 114 reconstructs the filtered out resolution from

15   the decoded video stream and provides a video data stream having the same resolution as the high-resolution input. However, because of the filtering and the losses resulting from the encoding and decoding, loss of information is present in the reconstructed stream. The loss is determined in the subtraction circuit 106 by subtracting the reconstructed high-resolution stream from the original, unmodified high-resolution stream. The output of the subtraction

20   circuit 106 is fed to an enhancement encoder 116 which outputs a reasonable quality enhancement stream.

Although these known layered compression schemes can be made to work quite well for progressive video, these schemes do not work well with video sent using interlaced SDTV standards. SDTV standards normally work well with interlaced video. For HDTV standards

25   both interlace and progressive HDTV standards are used. Although the known layered compression schemes work for movies, e.g., SD/HD DVD's, the known schemes do not provide a sufficient solution for interlace SDTV and HDTV.

SUMMARY OF THE INVENTION

30   The invention overcomes the deficiencies of other known layered compression schemes by introducing de-interlacers and re-interlacers into a layered compression scheme.

According to one embodiment of the invention, a method and an apparatus for efficiently performing spatial scalable compression of video information captured in a plurality of frames including an encoder for encoding and outputting the captured video

frames into a compressed data stream is disclosed. A base encoder for encoding an interlaced bitstream having a relatively lower pixel resolution. A spatial enhancement encoder for encoding a differential between a de-interlaced local decoder output from the base layer and an input signal.

5   According to another embodiment of the invention, a method and apparatus for encoding an input video stream is disclosed. An interlaced video stream is created from the input video stream. The interlaced stream is encoded to produce a base stream. The base stream is de-interlaced, decoded and optionally upconverted to produce a reconstructed video stream. The reconstructed video stream is subtracted from the input video stream to produce

10   a first residual stream. The resulting residual stream is encoded and outputted as an intermediate enhancement stream. The intermediate enhancement stream is temporal subsampled to produce a spatial enhancement stream.

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments described hereafter.

15

BRIEF DESCRIPTION OF THE DRAWINGS

          The invention will now be described, by way of example, with reference to the accompanying drawings, wherein:

          Figure 1 is a block diagram representing a known layered video encoder;

20   Figure 2 is a block diagram of a layered video encoder according to one embodiment of the invention;

          Figure 3 is a block diagram of a layered video decoder according to one embodiment of the invention;

          Figure 4 is a block diagram of a layered video encoder according to one

25   embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

          Figure 2 is a block diagram of a layered video encoder according to one embodiment of the invention. A high-resolution video stream 202 is inputted into a de-

30   interlacer 204. The de-interlacer 204 de-interlaces the input stream 202 and outputs a non-interlaced progressive signal composed of single frames. The non-interlaced signal is then downsampled by an optional downsampling unit 206. The decoupled video stream is then split by a splitter 208, whereby the video stream is sent to a second low pass filter/downsampling unit 210 and a subtraction unit 222. The low pass filter or

4

downsampling unit 210 reduces the resolution of the video stream, which is then fed to an interlacer 212. The interlacer 212 re-interlaces the video signal and then feeds the output to a base encoder 214. The base encoder 214 encodes the downsampled video stream in a known manner and outputs a base stream 216. In this embodiment, the base encoder 214 outputs a

5   local decoder output to a de-interlacer 218, which de-interlaces the output signal and provides a de-interlaced output signal to an upconverting unit 220. The upconverting unit 220 reconstructs the filtered out resolution from the local decoded video stream and provides a reconstructed video stream having basically the same resolution format as the high-resolution input video stream in a known manner. Alternatively, the base encoder 214 may output an

10   encoded output to the upconverting unit 220, wherein either a separate decoder (not illustrated) or a decoder provided in the upconverting unit 220 will have to first decode the encoded signal before it is upconverted.

The reconstructed video stream from the upconverting unit 220 and the high-resolution input video stream are inputted into the subtraction unit 222. The subtraction unit

15   222 subtracts the reconstructed video stream from the input video stream to produce a residual stream. The residual stream is then encoded by an enhancement encoder 224 to produce an intermediate enhancement stream 226. The intermediate enhancement stream is supplied to the temporal subsampling unit 242 which subsamples the intermediate enhancement stream to produce a spatial enhancement stream 244.

20   The encoder 214 also supplies the local decoder output to an addition unit 246, which combines the local base decoder output to a local enhancement decoder output from the enhancement encoder 224. The combined local decoder output is supplied to a splitter 230, which supplies the combined local decoder output to a temporal subsampling unit 232 and an evaluation unit 236. The temporal subsampling unit 232 performs the same temporal

25   subsampling as the encoder 214 performs on the original video input. The result is a 30 Hz signal. This reduced signal is fed to a motion compensated temporal interpolation unit 234, that is embodied in this example as a natural motion estimator. The motion compensated temporal interpolation unit 234 performs an upconversion from 30 Hz to 60 Hz by estimating additional frames. The motion compensated temporal interpolation unit 234 performs the

30   same upconversion as later the decoder will perform when decoding the coded data stream. Any motion estimation method can be employed according to the invention. In particular, goods results can be obtained with motion estimation based on natural or true motion estimation as used in for example frame rate conversion methods. A very cost efficient implementation is for example three-dimensional recursive search (3DRS) which is suitable

for consumer applications, see for example U.S. Patents 5,072,293, 5,148,269, and
5,212,548. The motion-vectors estimated using 3DRS tend to be equal to the true motion,
and the motion-vector field inhibits a high degree of spatial and temporal consistency. Thus,
the vector inconsistency is not thresholded very often and consequently, the amount of
5   residual data transmitted is reduced compared to non-true motion estimations.

The upconverted signal 235 is sent to an evaluation unit 236. As mentioned
above, the evaluation unit is also supplied with the combined local decoder output from the
splitter 230. The evaluation unit 236 compares the interpolated frames as determined by the
motion compensated temporal interpolation unit 234 with the actual frames. From the
10   comparison, it is determined where the estimated frames differ from the actual frames.
Differences in the respective frames are evaluated, in case the differences meet certain
threshold values, the differential data is selected as residual data. The thresholds can, for
example, be related to how noticeable the differences are, such threshold criteria per se are
known in the art. In this example, the residual data is described in the form of meta blocks.
15   The residual data stream 237 in the form of meta blocks is then put into an encoder 238. The
encoder 238 encodes the residual stream 237 and produces a temporal enhancement stream
240.

Figure 3 illustrates an exemplary decoder section according to one
embodiment of the invention. In the decoder section, the base stream 216 is decoded in a
20   known manner by a decoder 302, and the spatial enhancement stream 244 is decoded in a
known manner by a decoder 300. The decoded base stream is then de-interlaced by a de-
interlacing unit 306. The de-interlaced stream is then optionally upsampled in the
upsampling unit 308. The upsampled stream is then temporal subsampled by the temporal
subsampling unit 310. The subsampled stream is then combined with the decoded spatial
25   enhancement stream in the addition unit 312. The combined signal is then interpolated by a
motion compensating temporal interpolation unit 314. The temporal enhancement stream
240 is decoded in a known manner by a decoder 304. A combination unit 316 combines the
decoded temporal enhancement stream, the interpolated stream and the upsampled stream to
produce a decoder output.

30   Figure 4 illustrates an encoder according to another embodiment of the
invention. In this embodiment, a picture analyzer 404 has been added to the encoder
illustrated in Figure 2 to provide dynamic resolution control. A splitter 402 splits the high-
resolution input video stream 202, whereby the input video stream 202 is sent to the
subtraction unit 222 and the picture analyzer 404. In addition, the reconstructed video stream

from the upconverting unit 220 is also inputted into the picture analyzer 404 and the subtraction unit 222. The picture analyzer 404 analyzes the frames of the input stream and/or the frames of the reconstructed video stream and produces a numerical gain value of the content of each pixel or group of pixels in each frame of the video stream. The numerical

5   gain value is comprised of the location of the pixel or group of pixels given by, for example, the x,y coordinates of the pixel or group of pixels in a frame, the frame number, and a gain value. When the pixel or group of pixels has a lot of detail, the gain value moves toward a maximum value of "1". Likewise, when the pixel or group of pixels does not have much detail, the gain value moves toward a minimum value of "0". Several examples of detail

10 criteria for the picture analyzer are described below, but the invention is not limited to these examples. First, the picture analyzer can analyze the local spread around the pixel versus the average pixel spread over the whole frame. The picture analyzer could also analyze the edge level, e.g., abs of   -1-1-1

                                             -1 8-1

15   -1-1-1

per pixel divided over average value over whole frame.

         The gain values for varying degrees of detail can be predetermined and stored in a look-up table for recall once the level of detail for each pixel or group of pixels is determined.

20   As mentioned above, the reconstructed video stream and the high-resolution input video stream are inputted into the subtraction unit 222. The subtraction unit 222 subtracts the reconstructed video stream from the input video stream to produce a residual stream. The gain values from the picture analyzer 404 are sent to a multiplier 406 which is used to control the attenuation of the residual stream. In an alternative embodiment, the

25 picture analyzer 404 can be removed from the system and predetermined gain values can be loaded into the multiplier 406. The effect of multiplying the residual stream by the gain values is that a kind of filtering takes place for areas of each frame that have little detail. In such areas, normally a lot of bits would have to be spent on mostly irrelevant little details or noise. But by multiplying the residual stream by gain values which move toward zero for

30 areas of little or no detail, these bits can be removed from the residual stream before being encoded in the enhancement encoder 224. Likewise, the multipler will move toward one for edges and/or text areas and only those areas will be encoded . The effect on normal pictures can be a large saving on bits. Although the quality of the video will be effected somewhat, in

7

relation to the savings of the bitrate, this is a good compromise especially when compared to normal compression techniques at the same overall bitrate.

It will be understood that the different embodiments of the invention are not limited to the exact order of the above-described steps as the timing of some steps can be interchanged without affecting the overall operation of the invention. Furthermore, the term "comprising" does not exclude other elements or steps, the terms "a" and "an" do not exclude a plurality and a single processor or other unit may fulfill the functions of several of the units or circuits recited in the claims.